DRAFT

The Development of a DNA Procedure for the Forensic Identification of Caviar.

January 5, 2000

S. R. Fain, J. P. LeMay, J. Shafer, R. M. Hoesch and B. H. Hamlin

National Fish and Wildlife Forensics Laboratory, 1490 East Main Street, Ashland, OR, 97520, USA.
Tel: (541) 482-4191
FAX: (541) 482-4989
email: steve_fain@fws.gov

mitochondrial DNA, sequence analysis, species identification, sturgeon, caviar, forensic analysis

Abstract

DRAFT

This paper describes a method for identifying the species source of sturgeon/paddlefish caviar in international trade. The method utilizes phylogenetic analysis and bootstrap resampling of DNA sequence obtained from a 270 base pair segment of the mitochondrial cytochrome *b* gene. Maximum parsimony analysis of 246 type standards of 19 species obtained a phylogeny of 3 principle clades that is in significant agreement with expectations from sturgeon biogeography, cytogenetics and molecular systematics. The identification method positively identifies 15 of the world's 27 sturgeon and paddlefish species. Taxa that systematists have in the past regarded as conspecific, but now describe as distinct species (i.e., *Acipenser persicus – A. gueldenstaedtii*; *A. medirostris – A. mikadoi*; and *Scaphirhynchus platorhynchus – S. suttkusi*) were indistinguishable by this method.

We examined 907 different Beluga, Oscetra, Sevruga and American caviars as well as different sturgeon meats from North American, European and Caspian caviar retailers and from US imports submitted to the laboratory for verification of species origin persuant to CITES trade regulations. Fourteen species were identified: *Huso huso* (Beluga), *H. dauricus* (Kaluga sturgeon), *A. baerii* (Siberian sturgeon), *A. fulvescens* (Lake sturgeon), *A. gueldenstaedtii* (Russian sturgeon), *A. nudiventris* (Ship sturgeon), *A. oxyrhynchus* (Atlantic sturgeon), *A. persicus* (Persian sturgeon), *A. stellatus* (Stellate sturgeon), *A. shrenckii* (Amur sturgeon) *A. transmontanus* (White sturgeon), *Scaphirhynchus* spp. (likely Shovelnose sturgeon *S. platorynchus*), *Polyodon spathula* (American paddlefish) and a fourteenth unknown species with affinities to the North Pacific clade (likely *A. sinensis* or *A. dabryanus*). The method verified that 90% of US caviar imports are in compliance with the species origin listed on accompanying permits.

Introduction

The sturgeon (order Acipenseriformes) are primitive bony fish that evolved some 250 million years ago. The order is comprised of twenty-seven species in all: seventeen in the genus *Acipenser*, two in the genus *Huso*, three in the genus *Scaphirhynchus*, three in the genus *Pseudoscaphirhynchus* and two monotypic paddlefish genera - *Psepharus* and *Polyodon*. Adult fish range in size from just less than a meter to over five meters in length. Sturgeon have nearly been exterminated throughout their range due to intensive commercial fishing for caviar, and habitat perturbation in the last century (Birstein 1993). There are five sturgeon species in the Caspian Sea (i.e., *Huso huso* (Beluga sturgeon), *Acipenser gueldenstaedtii* (Russian sturgeon), *Acipenser stellatus* (Stellate sturgeon), *Acipenser persicus* (Persian sturgeon) and *Acipenser nudiventris* (Ship sturgeon). Of these, Beluga, Russian, Stellate and Persian sturgeon produce more than 90% of the caviar in world trade. The source countries of this production are the Russian Federation, Azerbaijan, Iran, Turkmenistan and Kazakhstan.

The precarious status of Caspian Sea sturgeon has resulted in the listing of all sturgeon species on the Convention on International Trade in Endangered Species (CITES) Appendix II. It has also initiated a shift in the world caviar market to other species, particularly North American White sturgeon and Paddlefish. A responsibility of the National Fish and Wildlife Forensic Laboratory (NFWFL) is to provide analytical support for the enforcement of the CITES concerning foreign or domestic trade in sturgeon products to facilitate sturgeon conservation.

Repeated efforts by the NFWFL to establish contacts in the Caspian and Black Sea regions indicated that it would be difficult if not impossible to obtain comprehensive sampling of geographically distinct populations of Eurasian sturgeon. This situation generated concern over the possibility that intraspecific variation would not be realistically represented in our reference base. Consequently, a conserved region of the mitochondrial cytochrome *b* gene (Irwin et al 1991) was chosen as a means to identify sturgeon species relevant to the caviar trade.

The cytochrome *b* gene has been used extensively as a genetic marker for vertebrate species distinction and, as a result, its evolutionary dynamics are well understood. The 5'end of cytochrome *b* has been shown to exhibit significant interspecific variation in sturgeon (Schill & Walker 1994, Birstein et al 1997) and other fish species (e.g., shark: Martin et al 1992; rockfish: Rocha-Olivares et al 1999) and, intraspecific sequence divergence has seldom exceeded a few percent (e.g., tuna: 1.0-2.0%, Bartlett & Davidson 1991; cod: 0-1.1%, Carr & Marshall 1992; stickleback: 0.4-3.1%, Orti et al 1994; rainbow fish: 0.5-3.6%, Zhu et al 1994).

To further reduce the potential effect of intraspecific variation on the identification of sturgeon species in trade, phylogenetically-informative sites were selected as species identifying characters. It was hypothesized that, especially within the more recently diverged species of *Acipenser*, synapomorphs would likely be "fixed" within species whereas autapomorphic-variable sites would be more likely to exhibit intraspecies variation. As such, the combined specificity of parsimony analysis on entire sequences, and on strings of phylogenetically

4

informative characters (haplotypes), would discriminate between sturgeon/paddlefish species and not be confounded by the variation extant within individual taxa.

This hypothesis was tested by first characterizing sequence variation in a 270 bp portion of the mtDNA cytochrome *b* gene (amino acid positions 36-125; nucleotide positions 108-378; Brown et al 1989) in 247 individual type standards from nineteen sturgeon and paddlefish species. Conclusions drawn from these data were then extended to similar characterization of 907 different test caviars and meats. We then combined the data sets and compared the phylogenetic identifications derived from parsimony analysis of total variation, and of informative sites only.

Materials & Methods

Sampling design

The published cytochrome *b* sequences of ten sturgeon and paddlefish species were used for primer design. These selections included representatives from each of the two families, three subfamilies and three tribes of the Acipenseriformes (Fig. 1). Both Eurasian and North American representatives were included in the selection. The GenBank/EMBL accession numbers and references for these sequences are listed in Table 1.

The 247 sturgeon/paddlefish standards used for the development of type sequences are listed in Table 2 along with the number of individuals characterized, and the source of each sample. Assessments of intraspecies variation in the targeted cytochrome *b* region included comparisons of: 17 samples from three populations of *Scaphirhynchus albus*, 79 samples from seven populations of *S. platorhynchus* and 3 samples from one population of *S. suttkusi*.

Within *Acipenser*, 24 samples from two populations of *A. transmontanus*; 10 samples from two populations of *A. medirostris*; and 7 samples of *Acipenser oxyrinchus oxyrinchus*, and 11 samples of *A. o. desotoi* were compared (Table 3). The type sequences from multiple accessions (range = 2 to 23, Table 2) of the thirteen remaining species were also examined for intraspecies variation.

Test sequences were obtained from 94 market-survey caviars purchased from 30 caviar retailers in North America, Europe and the Caspian Region (Table 4). In addition, 813 caviars from US imports were analyzed, pursuant to CITES trade regulations. Total cellular DNA was prepared from muscle tissue, whole blood or single eggs after the method of Boom et al (1990).

Cytochrome *b Sequences*.

The oligonucleotide primers used in this study for amplification and sequencing are listed in Table 5 and were suggested by (Kocher et al 1989), or designed to correspond to conserved regions of the cytochrome *b* sequences of sturgeon and paddlefish (Table 1).

Initially, sequences from a 423 base pair segment of the cytochrome *b* gene (amino acid positions 1-142) were obtained by the polymerase chain reaction (PCR) with primers L14769-13 & H15149 (Table 5). In subsequent applications of the method, primers L14851-13 & H14997 were used to amplify a segment corresponding to the internal 314 base pair portion of the original amplicon. The PCR amplifications were performed in 50 µL reaction volumes containing about 100 ng of template DNA (1-10 µL extract volume), 20 mM Tris (pH 8.4), 50 mM KCl, 2.0 mM MgCl$_2$, 200 µM concentration for each nucleotide, 1 µM concentration for each primer, Bovine Serum Albumin (160 µg/mL) and 2.5 units of Taq DNA polymerase (Life

Technologies). Amplification was accomplished in a programmable heating block (Model 9600; Perkin Elmer) and the amplification profile included 30 cycles of denaturing for 1 minute at 95°C, primer annealing for 45 seconds at 55°C, and extension for 45 seconds at 72°C. Smaller portions (154-315 bp) of the original amplicon were amplified from the DNAs of degraded caviars with internal primers (Fig. 1; Table 5) and the same cycling profile.

The amplification products were purified of unincorporated nucleotides and primers in dyalizing spin columns (Microcon-30, Millipore), brought to a final volume of 20 µL in sterile water and sequenced with the ABI PRISM BigDye Terminator Cycle Sequencing protocol (Perkin Elmer, 1998). Sequences were characterized on an automated DNA sequence analyzer (Model 377, Perkin Elmer). The region of the cytochrome *b* gene compared in the test for caviar species identification was 270 base pairs in length (amino acid positions 36-125; nucleotide sites 108-378)

Data analysis

The cytochrome *b* sequences of the type and test samples were aligned with the DNA Sequence Editor SeqEd (ver. 1.00A, Applied Biosystems, 1990). The PHYLIP (Phylogeny Inference Package) version 3.4 (Felsenstein, 1993), Maximum parsimony (DNAPARS) was used to estimate phylogenetic relationships among the nineteen sturgeon and paddlefish type sequences in Table 1, as well as between the type sequences and caviar test sequences. Phylogeny reconstructions of type sequences were tested with 500 bootstrap resamplings and heuristic searches (DNABOOT). Species identifications of individual test sequences were determined with 200 bootstrap resamplings. Phylogeny reconstructions were portrayed as majority-rule consensus trees (CONSENSE).

Results

Cytochrome b sequence evolution

Alignments of 270 bp of the 5' end of the cytochrome *b* gene identified a total of 76 variable sites that defined 24 unique sequences in 247 type standards of 19 sturgeon/paddlefish species. The ratio of variable sites among codon positions was 1:0:5. The ratio of transition to transversion substitutions was 5.6/1. Base substitutions produced 9 amino acid changes in the transcribed molecule, 4 of which were conservative. All *Acipenser* species in the Ponto-Caspian clade shared a synapomorphic valine at amino acid position 42 of the molecule. All of the *A. baerii* type standards had an autapomorphic isoleucine residue at position 66 of the molecule, whereas all of the other sturgeon/paddlefish analyzed had a valine at this position.

Interspecies variation among type standards

Inter-family (Polyodontidae v. Acipenseridae) sequence divergence (uncorrected) ranged from 11.5-15.2%; inter-subfamily (Husinae v. Acipenserinae) divergence ranged from 2.6-9.3% and the inter-tribe (Scaphirhynchini v. Acipenserini) divergence range was 7.0-11.1%. Sequence divergence between species of *Acipenser* ranged from 0-10.4%. Both of the *A. persicus* standards and eight of the ten *A. gueldenstaedtii* standards were identical. All of the *A. medirostris* and *A. mikadoi* sequences were identical. The *A. transmontanus* and *A. shrenckii* type sequences were distinguished at just one site (0.4% sequence divergence).

Intraspecies variation among type standards

The type standards of *Scaphirhynchus albus* and *S. platorhynchus* exhibited the same three sequences found among test caviars (Tables 3 & 6). All three were present together in two populations of *S. platorhynchus* whereas two type sequences were found together in two

populations of *S. albus*, and single populations of *S. platorhynchus* and *S. suttkusi*. The most divergent types in these species differed by 3 substitutions (1.1% sequence divergence), differing from the most common type (SplF11159 was exhibited by 86% of *S. platorhynchus*) by 1-2 substitutions (0.4-0.8% sequence divergence). Although intraspecies variation was evident in all three species, it was less than 1.1%, and it was not fixed within either species. No intraspecies variation was detected in the target region of cytochrome *b* for *A. transmontanus* - all 24 type sequences were identical. The 10 *A. medirostris* type standards were also identical to each other; as were all 18 *A. oxyrinchus oxyrinchus* and *A. o. desotoi* standards (Table 3).

In addition, the type standards for *A. ruthenus* and *A. stellatus* exhibited 3 unique sequences respectively (0.4-0.8% sequence divergence), and the *A. gueldenstaedtii*, *A. baerii* and *Polyodon spathula* type standards each exhibited 2 unique sequences (0.4% sequence divergence). However, no intraspecific variation was detected in the type sequences from multiple accessions (Table 2) of *A. brevirostrum, A. fulvescens, A. mikadoi, A. nudiventris, A. persicus, A. shrenckii, Huso huso* and *H. dauricus*.

Phylogenetic relationships of type standards

Comparisons of the type sequences revealed a total of 29 phylogenetically informative sites, 21 of which are synapomorphs (Table 6). Although informative, site #72 is a plesiomorphic character among *Huso huso*, site #117 is plesiomorphic among *Scaphirhynchus* spp., site # 132 is plesiomorphic among *A. transmontanus*, site #175 is plesiomorphic among *A. oxyrinchus,* and site #255 is effectively plesiomorphic among *Scaphirhynchus* spp. (i.e., the plesiomorphic character state is exhibited in only one variant of *A. stellatus*). Phylogenetically

informative sites #31, #90, #198 and #255 exhibited intraspecific variation in *Scaphirhynchus*,

Polyodon, Scaphirhynchus and *A. stellatus* respectively.

The phylogenetic relationships estimated from the type sequences (Fig. 2) stand in

substantial agreement with other published phylogenies based on biogeographic (Artyukhin

1995), cytogenetic (Birstein et al 1997) and molecular (Birstein & DeSalle 1998) assessments.

However, there are points of disagreement between the molecular phylogeny described here

and that proposed by Birstein and DeSalle (1998). Although the genus *Huso* is contained

within *Acipenser*, it is not monophyletic. *Huso huso* is basal to the Ponto-Caspian clade and

not most similar to *A. ruthenus*, while *Huso dauricus* is a part of the North Pacific species

clade, together with *A. medirostris*, *A. mikadoi*, *A. transmontanus* and *A. shrenckii*. In

addition, we find that the species *A. medirostris* and *A. mikadoi* are identical

Determining the species source of test samples   .

We obtained 270 bp of DNA sequence from the 5' end of cytochrome *b* from 976 test

caviars and meat items. A total of 79 variable sites were identified (additional sites were #3,

#87 and #187) defining 49 unique sequences, 25 of which were exclusive to the test samples

(Table 6).

Comparisons of the 24 unique cytochrome *b* type sequences with the 25 unique test

sequences resulted in the phylogenetic relationships depicted in Figure 3. The bootstrapped

phylogeny unambiguously (>60%) grouped 22 of the 25 test sequences with type sequences.

Thirteen species were positively identified: *Huso huso* (Beluga), *H. dauricus* (Kaluga sturgeon),

Acipenser baerii (Siberian sturgeon), *A. fulvescens* (Lake sturgeon), *A. gueldenstaedtii*

(Russian sturgeon), *A. nudiventris* (Ship sturgeon), *A. oxyrhynchus* (Atlantic sturgeon), *A.*

persicus (Persian sturgeon), *A. stellatus* (Stellate sturgeon), *A. shrenckii* (Amur sturgeon) *A. transmontanus* (White sturgeon), *Scaphirhynchus* spp. (likely Shovelnose sturgeon *S. platorynchus*) and *Polyodon spathula* (American paddlefish).

Test sequence G102011U7, was placed in the North Pacific clade (64% bootstrap value by MP; 78% bootstrap value by NJ) basal to *A. shrenckii* (AshF40705) and *A. transmontanus*, but could not be identified to species source. G102011U7 differed from AshF40705 by 8 substitutions (3.0% sequence divergence) and from AtrF10246 by 9 substitutions (3.3% sequence divergence). Although placed within the *A. gueldenstaedtii* species group by majority-rule consensus, the species source of test sequences 8487L2AU3 (Fig. 3h) and 8663L6BU28 (Fig. 3g) were not definitively supported by bootstrap analysis (>59%). Similarly, test sequence 8513L11U34 was placed within the *A. baerii* species group by majority-rule consensus (Fig. 3k), but its species source was not definitively supported by bootstrap analysis.

We observed the highest amount of intraspecific variation among the type standards and caviar test sequences grouped by the analysis with *A. gueldenstaedtii* (10 sequences) and *A. stellatus* (8 sequences). A total of 9 nucleotide sites varied among *A. gueldenstaedtii* test sequences: 7 third position and 2 first position, while 7 nucleotide sites varied among *A. stellatus*: 6 third position and 1 first position. Only the most commonly observed test sequence in each species group was represented in the reference base of type sequences- 67% of the 308 *A. gueldenstaedtii* test sequences were of sequence type AguG20874, 88% of the 223 *A. stellatus* test sequences were of sequence type AstG20904. The maximum divergence between haplotypes in these species was 4 substitutions (1.5%). Just one of the variable

nucleotide sites found among *A. stellatus* test sequences occurred at a phylogenetically informative site (Table 6).

In addition, four new test sequence variants were observed among caviars assigned to the *A. baerii* group, and three were observed among caviars assigned to the *Huso huso* group. A total of 5 nucleotide sites varied among *A. baerii* test sequences, and 3 nucleotide sites varied among *Huso huso*. Again, only the most commonly observed test sequence in each species group was represented in the reference base of type sequences - 86% of the 98 *A. baerii* test sequences were of sequence type AbaG41168, 86% of the 154 *H. huso* test sequences were of sequence type HhuG20913. The most divergent test sequences in the *A. baerii* group differed by 3 substitutions (1.1% sequence divergence), the test sequences in the *H. huso* group differed by a maximum of 2 substitutions (0.8% sequence divergence). One of the variable nucleotide sites found among the *A. baerii* test sequences occurred at a phylogenetically informative site (Table 6). None of the variable sites observed in the *Huso huso* test sequences occurred at phylogenetically informative sites.

New sequence variants were also observed among the test caviars assigned to the *A. nudiventris* and *Polyodon spathula* groups. The test sequence that grouped with *A. nudiventris*, B1AU13, differed from the type sequence AnuG41203 by 1 substitution (0.4% sequence divergence) at phylogenetically informative site 19 (Table 6). The AnuG41203 type sequence was observed in 94% of the 35 test sequences assigned to *A. nudiventris*. The test sequence that was grouped with *Polyodon spathula*, F40232U42, differed from the type sequence PspGBM64910 by 1 substitution (0.4% sequence divergence). The substitution did

not occur at a phylogenetically informative site (Table 6). The PspGBM64910 type sequence was observed in 72% of the 25 test sequences assigned to *Polyodon spathula*.

Discussion

DNA sequence analysis provided conclusive forensic species identifications from the basic unit of evidence, a single egg, as fully 719 of 907 caviar test sequences (79%) were identical to the sequences of type standards. DNA sequence analysis was robust in the face of commercial caviar processing and trade practices - it recovered unambiguous sequence types in 75% of over a thousand tests. Phylogenetic analysis of partial cytochrome *b* sequences identified markers in sturgeon and paddlefish that exhibited sufficient interspecies variation without confounding intraspecific variation: 79 variable sites describing 49 unique sequence types resolved to 29 species specific haplotypes of 29 phylogenetically informative sites each. In addition, statistical statements of confidence from bootstrap analysis were placed on species identifications drawn from the phylogenetic interpretation of these DNA sequence markers. Although four sturgeon/paddlefish species examined by this method were indistinguishable, fifteen were clearly identified – conspecific type and test sequences were grouped with >59% bootstrap support

Consistant with characterizations of other fish species, we identified low levels of intraspecific cytochrome *b* diversity among sturgeon/paddlefish. Maximum intraspecies sequence divergence was 1.5% among the type and test sequences comprising the *A. gueldenstaedtii* and *A. stellatus* species groups, followed by 1.1% sequence divergence among the *A. baerii* type and test sequences, and only 0.8% sequence divergence among *H.*

huso type and test sequences. Moreover, tests designed to estimate intraspecific variation in populations of *A. transmontanus, A. medirostris* and the subspecies *A. oxyrinchus oxyrinchus* and *A. o. desotoi* failed to detect any variation at all. In what might be considered a contradiction, however, representatives of the three "species" groups *A. persicus – A. gueldenstaedtii*; *A. medirostris – A. mikadoi*; and *Scaphirhynchus albus – S. platorhynchus – S. suttkusi* were also identical. It should be noted, however, that systematists have regarded these groups as conspecific in the past (Bailey & Cross 1954, Birstein & Bemis 1997). These results confirm the appropriateness of the selected region of the cytochrome *b* gene for the forensic species identification of sturgeon/paddlefish caviar. They also support the use of parsimony analysis and phylogenetically informative characters to discriminate between sturgeon/paddlefish species.

Phylogenetic characterizations of sturgeon/paddlefish type standards.

The phylogenetic relationships and geographic assortment of the species clades obtained in this study are, in the main, consistent with other published phylogenies based on biogeography (Artyukhin 1995), cytogenetics (Birstein et al 1997) and molecular systematics (Birstein & DeSalle 1998). However, these data are not intended to definitively describe the branching order that has occurred within the family and subfamilies of sturgeon/paddlefish. They are intended for forensic application and, as such, these data provide strong support at the terminal nodes defining the individual species of *Polyodon*, *Scaphirhynchus*, *Acipenser* and *Huso*.

Although similar, the phylogenetic relationships obtained in this study are not in complete agreement with those of Birstein & DeSalle (1998). The principal disagreements are that *Huso* is polyphyletic within *Acipenser*. *Huso huso* is not most closely related to *A. ruthenus*, it is the sister species to *A. stellatus*, *A. nudiventris*, *A. ruthenus*, *A. gueldenstaedtii* and *A. baerii*. In addition, *H. dauricus,* as part of the North Pacific clade along with *A. medirostris* and *A. mikadoi*, is basal to all other *Acipenser.* We also find *A. medirostris* and *A. mikadoi* to be identical. A partial explanation for these inconsistencies is the significant disagreement we have observed between the published sequences for *H. dauricus, A. ruthenus* and *A. medirostris* (Birstein et al 1997) and the type sequences for these species determined in this study. The sequences obtained in this study for *H. dauricus* and *A. ruthenus* (among others) were independently corroborated (Susan Homma & Hartmut Rehbein, personal communication). We obtained corroboration of our type sequence for *A. medirostris* by comparing multiple individuals from different localities. As a result, the published sequences for *H. dauricus, A. ruthenus* and *A. medirostris* (Birstein et al 1997) were not included in the phylogenetic analysis performed in this study.

Qualifications of molecular phylogenetics for forensic species identification.

In their influential work on the forensic identification of whales, Baker et al (1996) discuss how molecular phylogenetic species identification can be compromised by the use of incomplete type standard databases. If the species that an unknown test sequence originates from is not represented in the phylogenetic comparison, the test sequence can be incorrectly identified with the "most similar" species in the test – not with the actual species of identity. This

possibility is particularly relevant to the study at hand as type standards for *Acipenser dabryanus, A. sinensis, A. sturio, A. naccarii, Pseudoscaphirhynchus feldtschenkoi, P. hermanni, P. kaufmanni*, and *Psepharus gladius* were not included. Although *A. dabryanus* and *A. sinensis* have been the subject of caviar fisheries in the past, none of these eight species is currently considered to be sufficiently plentiful as to support commercial exploitation. In fact, *Pseudoscaphirhynchus* and *Psepharus* are virtually extinct (Birstein & Bemis 1997). In this situation, it is prudent to utilize all of the information that is available from comparisons of the test and type standard sequences (e.g., bootstrap value, sequence divergence, phylogenetically informative nucleotides or amino acids) as well as an appreciation of the general biogeographical concordance of the sturgeon/paddlefish phylogeny. An added precaution (Baker et al 1996) is to positively identify the species origin of a test sequence only when it groups within a set of type standards (see for examples Figs. 3b, 3e and 3r). Species identifications can be made when a test sequence is identified by significant bootstrap value as monophyletic with but basal to the node that defines a given type standard, but only with additional information.

As might be expected for ancient species such as the sturgeon/paddlefish, significant interspecific sequence variation was consistently observed. Inter-family (Polyodontidae v. Acipenseridae) sequence divergence ranged from 11.5-15.2%; inter-tribe (Scaphirhynchini v. Acipenserini) divergence range was 7.0-11.1%; and the sequence divergence between non-identical species of *Acipenser* and *Huso* ranged from 0.4-10.4%. The average interspecies sequence divergence was 5.8%, while the maximum intraspecific divergence was only 1.5%. As

these values are roughly concordant with phylogenetic distance, they are useful in informing the species identification process.

Transversions and amino acid substitutions have been used as phylogenetic markers to identify older divergences between related fish species (Meyer 1993). As transversion substitutions are prominent in our analysis of sturgeon/paddlefish, and most of them are fixed differences between species, they too can be used to further inform the species identification process. For example, both of the sites that vary among *A. ruthenus* type sequences are transitions, but there are two fixed transversion substitutions at sites #135 and #213 that distinguish *A. ruthenus* within *Acipenser*. Variants among *Polyodon spathula* type and test sequences exhibit 3 transitions, but there are five fixed transversion substitutions at sites #6, #114, #177, #222 and #261 that distinguish *Polyodon* from the Acipenseridae. In contrast, two of the 8 substitutions among *A. stellatus* variants are transversions. One is at site #12 in type sequence AstG20907, and a second is at site #213 of caviar test sequence 8378L1AU14 in the *A. stellatus* group. However, neither transversion substitution is fixed within *A. stellatus*.

The outcome of transversion substitutions is frequently amino acid replacement, i.e., they are often "non-silent" substitutions. Accordingly, we have observed that all *Acipenser* species in the Ponto-Caspian clade shared a synapomorphic valine at amino acid position 42 of the molecule, whereas none of those in the North Pacific clade did. All of the *A. baerii* type standards and caviar test sequences had an autapomorphic isoleucine residue at position 66 of the molecule that distinguished them within *Acipenser*.

We close with an example of how we combined molecular phylogenetic analysis with bootstrap resampling, sequence divergence, phylogenetically informative amino acid

polymorphisms and biogeography to infer the species origin of a test caviar from our type

standard database. In the present study, one test sample, G102011U7, was placed in the

North Pacific clade (65% bootstrap value) with *A. shrenckii* (AshF40705) and *A.*

transmontanus (Fig. 3v). However, G102011U7 could not be identified to species source

because its sequence differs from that of AshF40705 by 8 substitutions (3.0% sequence

divergence) and from AtrF10246 by 9 substitutions (3.3% sequence divergence). As these

divergence values are more than twice the maximum observed intraspecific divergence, and

because G102011U7 exhibits the amino acid valine polymorphism at position #42 in contrast to

the type standards from this clade, it is likely that U7 represents a 14th caviar trade species. The

biogeography of the North Pacific clade suggests that the likely species source is either

Acipenser dabryanus or *A. sinensis*.

DRAFT

Conclusions

The DNA sequence analysis procedure that we developed for the forensic identification of caviar was sensitive enough to provide a conclusive answer from the basic unit of evidence, a single egg. It was robust in the face of commercial caviar processing and trade practices (e.g., pasteurization, salting and various conditions of storage). It is increasingly difficult to obtain type standards of endangered species such as sturgeon/paddlefish in numbers sufficiently representative for the traditional identification methods of morphology and population genetic analysis. An alternative approach to the challenges of forensic species identification today is to change the markers that have been traditionally used. The information content of mitochondrial and nuclear DNA is enormous, providing copious options for marker selection. Phylogenetic analysis of these markers can predict those that will exhibit sufficient interspecies variation without confounding intraspecies variation. In addition, statistical statements of confidence can be placed on conclusions drawn from the phylogenetic interpretation of DNA sequence markers as to the species source of unknowns. This is an increasingly frequent request of the court.

References DRAFT

1. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and I.G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.

2. Artyukhin, E. N. 1995. On biogeography and relationships within the genus *Acipenser*. *Sturgeon Quarterly* 3:6-8.

3. Bailey, R.M. and F.B. Cross. 1954. River sturgeons of the American genus *Scaphirhynchus*: characters, distribution, and synonymy. *Papers Michigan Academy Science, Arts and Letters* 39:169-208.

4. Baker, C.S., Cipriano, F. and S.R. Palumbi. 1996. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. *Molecular Ecology* 5:671-685.

5. Bartlett, S.E. and W.S. Davidson. 1991. Identification of *Thunnus* Tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Canadian Journal of Fisheries and Aquatic Sciences* 48:309-317.

6. Beckenbach, A.T., Thomas, W.K. and H. Sohrabi. 1990. Intraspecific sequence variation in the mitochondrial genome of rainbow trout (*Oncorhynchus mykiss*). *Genome* 33:13-15.

7.   Birstein, V.J., Hanner,R. and R. DeSalle. 1997. Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. *Env. Biol. Fish.* 48:127-155.

8.   Birstein, V.J. and W.E. Bemis. 1997. How many species are there in the genus *Acipenser*? *Env. Biol. Fish.* 48:157-163.

9.   Birstein, V.J. and R. DeSalle. 1998. Molecular phylogeny of Acipenserinae. *Mol. Phylogenet. Evol.* 9:141-155.

10.   Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E. and J.van der Noordaa. 1990. Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*; 28 (3): 495-503.

11.   Bowen, B.W. and W.S. Grant. 1997. Phylogeography of the sardines (*Sardinops* spp.): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution* 51:1601-1610.

12.   Brown, J.R., Gilbert, T.L., Kowbel, D.J., O'Hara, P.J., Buroker, N.E., Beckenbach, A.T. and M.J. Smith. 1989. Nucleotide sequence of the apocytochrome B gene in white sturgeon mitochondrial DNA. *Nucleic Acids Res.* 17:4389-

13.   Carr, S.M. and H. D. Marshall. 1992. Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gaduas morhua*) by the polymerase chain reaction. *Can. J. Fish. Aquat. Sci.* 48:48-52.

14.   Carr, S.M., Snellen, A.J., Howse, K.A. and J.S. Wroblewski. 1995. Mitochondrial DNA sequence variation and genetic stock structure of Atlantic cod (*Gaduas morhua*) from bay and offshore locations on the Newfoundland continental shelf. *Mol. Ecol.* 4:79-88.

15.   Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.

16.   Felsenstein, J. 1993. Phylip (Phylogeny Inference Package). Seattle, WA Univ. of Washington

17.   Findeis, E.K. 1997. Osteology and phylogenetic interrelationships of sturgeons (Acipenseridae). *Env. Biol. Fish.* 48:73-126.

18.   Irwin, D.M., Kocher, T.D. and A.C. Wilson. 1991. Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol.* 32:128-144.

19.   Martin, A.P., Naylor, G.J.P. and S.R. Palumbi. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* 357:153-155.

20.   Meyer, A. 1993. Evolution of mitochondrial DNA in fishes. *In: "Biochemistry and Molecular Biology of Fishes".* (P.W. Hochachka and T.p. Mommsen, Eds.), pp. 1-38, Elsevier, Amsterdam.

21.   Normark, B.B., McCune, A.R. and R.G. Harrison. 1991. Phylogenetic relationships of neopterygian fishes, inferred from mitochondrial DNA sequences. *Mol. Biol. Evol.* 8:819-834.

22.   Orti, G., Bell, M.A., Reimchen, T.E. and A. Meyer. 1994. Global survey of mitochondrial DNA sequences in the threespine stickleback: Evidence for recent migrations. *Evolution* 48:608-622.

23.   Rocha-Olivares, A., Kimbrell, C.A., Eitner, B.J. and R.D. Vetter. 1999. Evolution of a mitochondrial cytochrome *b* gene sequence in the species-rich genus *Sebastes*

(Teleostei, Scorpaenidae) and its utility in testing the monophyly of the subgenus *Sebastomus*. *Mol. Phylogenet. Evol.* 11:426-440.

24.   Schill, W.B. and R.L.Walker. 1994. Phyogenetic relationships of sturgeon inferred from cytochrome *b* sequences. Unpublished Technical Report. .

25.   Zhu, D., Jamieson, B.G., Hugall, A and C. Moritz. 1994. Sequence evolution and phylogenetic signal in control-region and cytochrome *b* sequences of rainbow fishes (Melanotaeniidae). *Mol. Biol. Evol.* 11:672-683.